



## **Reevaluation of genetic variants previously associated with arrhythmogenic right ventricular cardiomyopathy integrating population-based cohorts and proteomics data**

Ye, Johan Z.; Delmar, Mario; Lundby, Alicia; Olesen, Morten S.

*Published in:*  
Clinical Genetics

*DOI:*  
[10.1111/cge.13621](https://doi.org/10.1111/cge.13621)

*Publication date:*  
2019

*Document version*  
Peer reviewed version

*Citation for published version (APA):*

Ye, J. Z., Delmar, M., Lundby, A., & Olesen, M. S. (2019). Reevaluation of genetic variants previously associated with arrhythmogenic right ventricular cardiomyopathy integrating population-based cohorts and proteomics data. *Clinical Genetics*, 96(6), 506-514. <https://doi.org/10.1111/cge.13621>

# Reevaluation of Genetic Variants Previously Associated with Arrhythmogenic Right Ventricular Cardiomyopathy Integrating Population-based Cohorts and Proteomics Data.

Johan Ziruo Ye<sup>1,2</sup>, Mario Delmar<sup>3</sup>, Alicia Lundby<sup>1,4,#</sup>, Morten S. Olesen<sup>1,5,#</sup>.

<sup>1</sup> *Department of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen N, Denmark.*

<sup>2</sup> *Department of Health Technology, Section for Bioinformatics, Technical University of Denmark, DTU, 2800 Kgs. Lyngby, Denmark*

<sup>3</sup> *The Leon H. Charney Division of Cardiology, New York University School of Medicine, New York, New York.*

<sup>4</sup> *Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen N, Denmark.*

<sup>5</sup> *Laboratory of Molecular Cardiology, Department of Cardiology, Heart Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark*

<sup>#</sup> *These authors contribute equally*

## Data Availability Statement

gnomAD data is publicly available at <https://gnomad.broadinstitute.org/downloads>

ClinVar data is publicly available at <https://www.ncbi.nlm.nih.gov/clinvar/>

ARVD/C Genetic Variant Database is publicly available at <https://molgenis136.gcc.rug.nl/>

HGMD data is publicly available at <http://www.hgmd.cf.ac.uk/>

## Abstract

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is one of the most common causes of sudden cardiac death in young people. Patients diagnosed with ARVC may experience increased likelihood of development of anxiety and depression, emphasizing the need for accurate diagnosis. To assist future

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/cge.13621

genetic diagnosis and avoidance of misdiagnosis, we evaluated the reported monogenic disease-causing variants in *ARVD/C Genetic Variants Database*, *Human Gene Mutation Database*, and *ClinVar*.

Within the aforementioned databases, 630 monogenic disease-causing variants from 18 genes were identified. In the *genome Aggregation Database*, 226 of these were identified; 68 of which were found at greater than expected prevalence. Furthermore, 37/226 genetic variants were identified amongst the 409,000 UK biobank participants, 23 were not associated with ARVC. Among the 14 remaining variants, 13 were previously found with greater than expected prevalence for a monogenic variant. Nevertheless, they were associated with serious cardiac phenotypes, suggesting that these 13 variants may be disease-modifiers of ARVC, rather than monogenic disease-causing. In summary, more than 10% of variants previously reported to cause ARVC were found unlikely to be associated with highly penetrant monogenic forms of ARVC. Notably, all variants in *OBSCN* and *MYBPC3* were found, making these unlikely to be monogenic causes of ARVC.

**Keywords (after Abstract):**

Arrhythmogenic right ventricular cardiomyopathy, gnomAD, Cardiac Proteomics, Genetic Variants, Monogenic, Disease-modifier, ARVC disease-modifier

**Introduction**

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is an inheritable cardiac muscle disorder, in which the myocardium is progressively replaced by fibrofatty tissue. Prevalence of the disease varies within regions, though in some areas is estimated to be responsible for as many as 20% of all cases of sudden cardiac death among people under 35 years of age<sup>1</sup>. The disease primarily involves the right ventricle, although left ventricular involvement occurs and in some cases predominates<sup>2</sup>. Additionally, exercise is known to accelerate the onset and/or progression of ARVC<sup>1</sup>. This emphasizes the need for early diagnosis and risk assessment of ARVC, particularly as life-threatening arrhythmias often occur during the concealed phase of the disease<sup>3</sup>. Unfortunately, early diagnosis of ARVC is particularly difficult. Current task force criteria<sup>4</sup> are used as standards for disease diagnosis and include genetic testing. However, as with other inheritable diseases, discerning the significance of genetic variants is particularly challenging. To assist in this challenge, we have intersected multiple publicly available databases to help critically evaluate the previously reported disease-causing genetic variants.

The genome Aggregation Database<sup>5</sup> (*gnomAD* r2.1.1) contains 123,136 unrelated exome sequences that are used to represent the general population in this study, hence do we expect occurrence of ARVC-causing genetic variants in this database (<http://gnomad.broadinstitute.org/>). *gnomAD* is the largest publicly available population data set, and the populations have been sub-grouped based on principal component analysis. This database may be used to evaluate the prevalence of genetic variants reported to cause ARVC in the general population and thus its likelihood of causing the disease. However, the *gnomAD* database lacks phenotype data. Thus, as an additional step we compared the identified disease-causing variants from *gnomAD* against PheWeb<sup>6</sup> (<http://pheweb.sph.umich.edu:5003/>), a phenome-wide association study based on 409,000 UK biobank participants. Finally, we integrated high-resolution proteomics data from each of the cardiac chambers of *Sus scrofa* hearts (n=3) in order to further support our categorization and evaluation of genes and their variants. These multi-step intersecting analyses allowed us to reduce the number of variants – and the number of genes - likely to be related to the disease. We propose that a continuous refinement through additional databases as they become available can serve as an added tool for diagnosis and assessment of risk in ARVC-affected population.

## Methods

### Databases searched for reportedly monogenic disease-causing variants

We evaluated pathogenic and likely pathogenic small deletions related to ARVC reported in the *Human Gene Mutation Database*<sup>7</sup> ([www.hgmd.cf.ac.uk](http://www.hgmd.cf.ac.uk) - HGMD Professional 2017.4, visited March 2018), ClinVar<sup>8</sup> (<https://www.ncbi.nlm.nih.gov/clinvar/> - visited March 2018) and in the *ARVD/C Genetic Variants Database*<sup>9</sup> (update from February 2015, visited March 2018 - <https://molgenis136.gcc.rug.nl/>). The variants in HGMD were queried using the phenotype: “arrhythmogenic right ventricular cardiomyopathy” and “cardiomyopathy, arrhythmogenic right ventricular”. ClinVar variants were identified by searching: “ARVC” OR “arrhythmogenic right ventricular”. Variants were only included if they were pathogenic/likely pathogenic and had the phenotype “ARVC” or “arrhythmogenic right ventricular”. Only ClinVar variants with multiple non-conflicting submitters or originating from: “research” or “literature only” were included. *JUP* and *DSP* variants reported to be recessively inherited in any database were removed.

### Variant prevalence analysis in *gnomAD*

The reportedly monogenic disease-causing variants were searched for in the 123,136 exomes in *gnomAD* (r2.1.1)<sup>5</sup>. Each gene's association to ARVC was ranked based on the amount of variants within it not found in *gnomAD*. Variants were only included if they had *gnomAD* filter status of "PASS", meaning the variant passed all *gnomAD* quality control filtering. Genotype prevalence was calculated based on the allele count, the amount of homozygous people with the variant, and coverage at the allele location. According to *American College of Medical Genetics and Genomics* (ACMG) and the *Association for Molecular Pathology* (AMP) guidelines for variant interpretation,<sup>10</sup> variant prevalence above disease prevalence is recognized as strong evidence for a variant being benign, regardless of its over-representation being in the general population or only in a subpopulation (ethnic group). Based on disease prevalence, a genotype prevalence above 1:2000 was considered greater than expected and was used as a cutoff for each individual variant. Using this data set, we can evaluate the prevalence of the reported ARVC variants. Additionally, all variants were scored using the *Combined Annotation Dependent Depletion* (CADD) scoring system to in-silico predict the deleteriousness of the ARVC variants<sup>11</sup>. Furthermore, we evaluated the number of variants found to be prevalent in the general population (in *gnomAD*) against the number of variants reported in each of the databases reporting ARVC-related variants.

#### **Phenome-wide association study data from 409,000 participants in UK biobank**

To support the prevalence analysis, phenotype evidence from PheWeb<sup>12</sup> (<http://pheweb.sph.umich.edu:5003/> - visited May 2018) was used. PheWeb is a phenome-wide association study conducted on the 409,000 Caucasian UK biobank participants. The following phenotypes were considered an indicator that the variant could predispose for ARVC: Any cardiomyopathy, "Heart failure," "Ventricular fibrillation," "myocarditis," "paroxysmal ventricular tachycardia," and "Cardiac arrest." A variant was also considered potentially predisposing for ARVC, if it displayed significant association with another pathogenic phenotype from the circulatory system after correction for multiple hypothesis testing.

#### **Proteomics analysis of protein abundance in cardiac chambers**

We used data from proteomics experiments of each of the four cardiac chambers, which were obtained from three adult pigs (*Sus scrofa*) (personal communication, Associate professor A. Lundby). Sample processing was performed in a similar way as recently published by Linscheid et al<sup>13</sup>. Briefly, cardiac tissue

Accepted Article

biopsies were collected and immediately snap frozen in liquid NO<sub>2</sub> and stored (–80 °C). Protein extraction of biopsies (2 mg per sample) was followed by digestion as described previously<sup>14,15</sup>. Desalted digested peptides were fractionated into 12 fractions by micro-flow reverse-phase ultrahigh pressure liquid chromatography. Following this, samples were separated on 15 cm fused-silica emitter columns (in a 1 h multi-step linear gradient). These were analyzed by online reversed-phase liquid chromatography coupled to a Q-Exactive Plus quadrupole Orbitrap tandem MS. Using the MaxQuant software, we processed the raw MS data. To identify proteins, we used a built-in Andromeda search engine containing human SwissProt protein entries. From these experiments, in total we measured 7206 proteins. Rank analysis was based on summed MS-based protein intensities. Protein abundance ratios between cardiac chambers were investigated by examining the median log<sub>2</sub> fold change difference between cardiac chambers from each pig. To investigate the relative abundance of proteins encoded by ARVC associated genes, proteins were ranked and plotted according to abundance using the median protein intensities from the right ventricle after iBAQ correction.

## Results

### Variants Identified in gnomAD

To evaluate variants that are reported to cause ARVC, we gathered variants claimed pathogenic or likely pathogenic in either ARVC-database<sup>16</sup>, ClinVar<sup>8</sup>, and HGMD<sup>7</sup>. A total of 630 unique genetic variants from 18 different genes reported ARVC-causing were identified, which can be found in supplementary file 1. The 630 disease associated variants were found across 18 different genes (*PKP2*, *DSP*, *DSG2*, *DSC2*, *LMNA*, *DES*, *JUP*, *RYR2*, *TTN*, *SCN5A*, *PLN*, *CTNNA3*, *CDH2*, *TMEM43*, *OBSCN*, *MYBPC3*, and *DMD*). A combined 87% of the genetic variants were found within the top 5 genes, *PKP2*, *DSP*, *DSG2*, *DSC2*, and *JUP*. In contrast, across the bottom 8 genes only 3% variants were identified (Table 1). The protein expression pattern between atria and ventricles were then examined (Figure 1). The five most common desmosomal genes have a general trend towards being expressed more highly within the ventricles (Figure 1a) and a small trend towards the right ventricle (Figure 1c). Meanwhile, the general trend of the remaining ARVC-associated genes is towards the atria (Figure 1b). Furthermore, to ensure that this discrepancy in expression pattern was not due to very low expression in one chamber, we plotted a rank plot of abundance levels within the right ventricle of the 18 genes. All except one protein were detected with above median intensity within the right ventricle. High expression does not exclude the variant from being benign, despite *MYBPC3* being

in the 95-percentile of protein expression within the ventricle, it is still expressed more so in the atria of the pig. Therefore, it is important to evaluate based on relative expression pattern and not absolute expression within a heart chamber. Using this information together with the number of variants within each gene identified within gnomAD, we categorize the genes depending on the confidence that the genes are indeed associated with ARVC (Table 1). From these 18 genes, the 630 genetic variants were next evaluated through their prevalence in the general population, and the phenotypes each variant is associated with.

### **Sixty seven variants with prevalence evidence for inaccurate variant interpretation**

From the total list of 630 unique genetic variants claimed to cause ARVC, 226 variants were found in the general population database, gnomAD. Identification of a genetic variant in gnomAD does not by itself preclude pathogenicity. According to ACMG and AMP guidelines<sup>10</sup> pathogenic variants with prevalence greater than the disease prevalence (genotype prevalence > 1:2000) in gnomAD is considered strong evidence for a variant being benign. From the 226 variants identified in gnomAD, 68 genetic variants were present at a frequency incompatible with causation of penetrant variants. Each of the 68 variants were found with prevalence above our conservative cutoff of 1:2000, in one or more subpopulation(s) of gnomAD. gnomAD's subpopulations are subsetted based on different population ethnicities (eg. African, Non-Finnish European, south Asian etc.). Furthermore, 36/68 variants had a prevalence above 1:2000 in the overall population in gnomAD, meaning each of these 36 variants should in theory alone account for all the expected the ARVC individuals within gnomAD. Therefore, these 68 variants reported to be monogenic causes of ARVC have strong evidence for being benign and thus for currently being inaccurately interpreted. All 630 variants were assessed in silico using a meta-score prediction tool, CADD<sup>11</sup>. CADD scores of the variants identified in gnomAD were compared to those not found using student's t-test, which resulted in a p-value 6.35e-8, indicating there was a significant difference in CADD scores between variants found in gnomAD, and those not found, supporting our gnomAD findings. When diving into these CADD scores we saw that the percentage of variants that scored in the top 0.1% most deleterious variants dropped considerably between variants not found in gnomAD (50.56%), those found in gnomAD with low prevalence (22.22%), and those found in gnomAD at prevalence incompatible with monogenic disease-causing variants (7.5%) (see figure 2).

### **Twenty four variants not found significantly associated to serious cardiac phenotypes.**

Following the evaluation of variant frequency, the phenotype associated with the genetic variants found in gnomAD were investigated via query in the PheWeb database. In PheWeb, 37 variants out of the 226 variants identified in gnomAD were identified. Out of these 37 variants, all but two variants (35) had previously been found to have a frequency incompatible with pathogenic variants in gnomAD, which indicated these 35 were benign. In PheWeb, 23 of these 37 variants showed no sign of being associated with any ARVC-like phenotypes supporting their evaluation based on their genotype prevalence (see table 2). However, 14 variants showed an association with ARVC-like phenotypes (nominal  $p < 0.05$ ), but only one was significant after correcting for multiple statistical testing. All associated ARVC-like phenotypes display large odds-ratios, indicating they have strong association to their phenotype(s). One variant (rs201856473) was of low prevalence (Genotype prevalence = 0.00013) in the gnomAD database and displayed association to ARVC-related phenotypes in UK biobank, indicating it is likely disease predisposing. Of the remaining 23 variants that did not display association to ARVC-related phenotypes in UK Biobank, 22 had previously been found with greater than expected prevalence in gnomAD. The single aforementioned variant (rs201564919) with gnomAD prevalence below our cutoff was found in PheWeb to show no association to serious cardiac phenotypes, which is considered strong evidence for benignity. Adding the single variant with gnomAD prevalence below our threshold that displayed no signs of being associated to ARVC in PheWeb, brings the total variants with evidence of being inaccurately classified to 69. However, 13 variants were found to be associated with serious cardiac phenotypes, conflicting with their evidence for benignity from the prevalence analysis.

#### ***ARVD/C Genetic Variants Database is the preferred database according to gnomAD analysis***

the three databases the variants are sourced from, HGMD displays the highest number of variants detected in gnomAD. ClinVar and the ARVC-database contained a similar percentage of variants also found at high levels in gnomAD; however, the ARVC-database includes more variants than the other databases (by a factor of 4) and as such, showed the best performance in terms of total number of variants reported after excluding those of high prevalence in the general population (see figure S1).

## **Discussion**

### **Thirteen inaccurately classified ARVC variants predispose for other serious cardiac phenotypes**



Over 1/3 of the monogenic disease-causing variants were identified within gnomAD (226/630), which is in agreement with a previous study by Hall et al, which focused only on variants from the ARVC database, identifying 117 out of 364 pathogenic variants<sup>17</sup>. The 226 variants with questionable ARVC association were evaluated using phenome-wide association study data from UK biobank (PheWeb). A total of 37 of these variants were found within the UK biobank data set. Approximately two thirds of these variants (23/37) were not associated with any phenotypes that may associate to ARVC, thus suggesting that they are not pathogenic (Table 2). The remaining 14 variants appeared to predispose to cardiac phenotypes. The variant with the lowest p-value association to its phenotype was variant rs147240502 ( $p = 0.000012$ ), however, the reported association in PheWeb was to the term “aortic aneurysm”, a phenotype not directly related to ARVC. Interestingly, 13/14 variants were found associated with one or more serious cardiac phenotype(s) in PheWeb, despite displaying greater than expected prevalence in the overall gnomAD population. Additionally, it is noteworthy that 11 of these variants were found with odds-ratios above 5 (see table 2). We therefore propose that these 13 variants are previously unreported disease modifiers of ARVC, as they are unlikely to be monogenic disease-causes of ARVC.

#### **Challenging three entire genes of their association to ARVC as sources of monogenic pathogenic variants**

One fourth (68/226) of the variants found in gnomAD displayed a genotype prevalence above cutoff in the overall gnomAD population, or in one of its subpopulations, thus challenging their current status as monogenic disease-causing variants. It is important to note that these 68 variants are found at a rate incompatible with monogenic disease-causing variants of ARVC. This does not exclude the variants from being disease-causing of other diseases. They are only classified as highly unlikely to be monogenic disease-causing of ARVC. Furthermore, we identified three genes (*OBSCN*, *MYBPC3*, and *DMD*) to be “low confidence” genes (Table 1) given that all the variants within these genes were found in gnomAD and due to their relative low expression pattern in the ventricles compared to atria (Figure 1). We expect the protein expression to trend towards being higher in the RV, since variants within these genes are reported to cause ARVC, which manifests itself primarily in the right ventricle. However, most of the medium and low-confidence genes are expressed more so in the atria. The percentage of variants found in gnomAD within the genes and their protein expression allows us to question the actual association of the three bottom genes (*OBSCN*, *MYBPC3*, and *DMD*) to ARVC. Not only did *OBSCN* and *MYBPC3* have all their variants identified in gnomAD, their two variants were both identified at an occurrence incompatible with

disease-causing variants. From the results of our study, we propose that these three genes ought to be evaluated further, to conclusively determine if variants within them are able to cause ARVC. Note that for *JUP*, a desmosomal protein, 19/24 of its variants identified in gnomAD and 2 of these had a high prevalence. Recessive JUP mutations have been associated to the recessive genetics disease, Naxos disease, a cardiocutaneous syndrome which may complicate interpretation if this was not reported clearly in the original publications.

### **Monogenic disease-causing variants' genotype prevalence highly unlikely to exceed 1:2000**

Our analysis retrospectively examines variants that were associated to patients with the diagnosis of ARVC. With our study, we seek to improve the clarity of interpretation of variants reported during genetic testing. We have intersected existing databases as a mean of exploring the likelihood of a reported variant in ARVC-related databases to be causative of disease. We used the prevalence of 1:2000 as the cutoff value in gnomAD. This is a very conservative approach, since even variants in the most common gene (*PKP2*) are only found in 10-45% of patients<sup>3</sup>. We based this cutoff on the disease prevalence in people of European descent, which is the largest subpopulation in gnomAD<sup>5</sup>. We reasoned that if the variant is disease-causing, then the disease prevalence should at least match the occurrence of the variant. By using a 1:2000 (rather than a 1:5000) cutoff, we conservatively estimated the likelihood of a variant being benign and not linked to disease. Our study also found that genes *OBSCN*, *MYBPC3*, and *DMD* are unlikely to be causative of the ARVC phenotype. Especially *OBSCN* given that the variant reported in these genes are prevalent above 1:2000 in the overall population of gnomAD.

### **Identification of Potential Disease-modifying Genetic Variants**

Our analytical methods were able to identify potential a substantial number of non-monogenic disease-causing genetic variants. Interestingly, some of these down-classified variants seem to act as disease-modifiers. For example, the variant with the second strongest association to its phenotype "(Other cardiomyopathies)" is *DSG2* variant rs372174546. This variant also exhibits a very large odds-ratio, and its CADD score tells us that it is among the top 10% most deleterious variants. Yet, this variant has been found to have a prevalence greater than expected, although the location does not have the greatest coverage (~25,000 allele reads), it is present in 3 times the expected amount in non-Finnish-Europeans. The greater than expected prevalence suggests that this variant is not a monogenic disease-causing variant, instead we

propose that this variant is a disease-modifier. On the other hand, the *DSC2* variant rs201856473 had a low in silico prediction score, indicating it could be a benign variant, yet it was found with the strongest association to its phenotype (“ventricular fibrillation and flutter”) of tested variants with odds-ratios greater than 5. This variant was also found with prevalence below our threshold in gnomAD. These two examples emphasize the importance of the clinical presentation of ARVC within the family, as well as other criteria, well established by the Task Force<sup>4</sup> need to be taken into account before a diagnosis and risk stratification is established.

The single variant from *OBSCN*, rs200362121, was found in greater than expected frequency, however, it was also found with association to “cardiomyopathies” in Pheweb. We propose this *OBSCN* variant could be a disease modifier rather than a monogenic disease-causing variant. The single variant from *MYBPC3* (rs771143409) is identified in gnomAD in a frequency incompatible with monogenic disease-causing variants. Thus, *OBSCN* and *MYBPC3* has their association to ARVC questioned, leaving very limited evidence that the gene is disease causing gene for ARVC.

### Limitations

The ACMG and AMP standards<sup>10</sup> are subject to potential bias due to categorization depending on individual assessments. We have not assessed whether one or more disease-causing variant subsides within each individual. Our study performed substantial computer prediction on variants from three separate databases, thus we did not take experimental data into account. Please see original references to view these data. We cannot rule out that some variants could increase susceptibility of ARVC under certain circumstances such as strenuous exercise, which has been shown to de-mask potential disease-causing variants<sup>18</sup>.

Genetic screening may assist identification of patients and family members, meaning misclassified genetic variants in patients may lead to erroneous risk stratification or misdiagnosis. This could have serious clinical consequences, emphasizing the importance of variants being reported causative being truly disease-causing.

### Conclusion

We have identified many variants previously considered monogenic causes of ARVC, and >25% were found in a large general population exome database. Furthermore, more than 10% of the genetic variants were

Accepted Article

found at a prevalence incompatible with disease causing variants according to *American College of Medical Genetics and Genomics* (ACMG) and the *Association for Molecular Pathology* (AMP) variant interpretation guidelines<sup>10</sup>. Included in these 10% were all variants reported from *OBSCN* and *MYBPC3* indicating that these genes do not need to be sequenced when attempting to identify the causal genetic variant. Interestingly, a focused analysis using over 400,000 UK biobank participants suggests that 13 genetic variants play a disease-modifying role rather than a monogenic disease-causing role for ARVC. We therefore, advise caution when incorporating genetic tools in the diagnosis of ARVC.

### Acknowledgements

*The study was supported by Support from The Novo Nordisk Foundation (NNF18OC0052844) and Support from The Danish Council for independent Research (DFF-4092-00045 and DFF-6110-00166) to AL.*

*The authors wish to thank the genome aggregation database (gnomAD) and the contributing projects in creating this resource for biomedical research. The authors would also like to thank the Human Gene Mutation Database (HGMD), ClinVar, and ARVD/C Genetic Variants Database for providing exome data during the years.*

**The authors have no conflicts of interest to declare.**

- Accepted Article
1. Thiene, G., Nava, A., Corrado, D. & Rossi, L. Right ventricular cardiomyopathy and sudden death in young people. *N. Engl. J. Med.* **318**, 129–133 (1988).
  2. Corrado, D., Basso, C. & Thiene, G. Spectrum of Clinicopathologic Manifestations of Arrhythmogenic Right Ventricular Cardiomyopathy / Dysplasia : A Multicenter Study. **30**, 1512–1520 (1997).
  3. Corrado, D., Link, M. S. & Calkins, H. Arrhythmogenic Right Ventricular Cardiomyopathy. *N. Engl. J. Med.* **376**, 61–72 (2017).
  4. Marcus, F. I. *et al.* Diagnosis of arrhythmogenic right ventricular cardiomyopathy/Dysplasia: Proposed modification of the task force criteria. *Circulation* **121**, 1533–1541 (2010).
  5. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
  6. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
  7. Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
  8. Landrum, M. J. *et al.* ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, 980–985 (2014).
  9. Lazzarini, E. *et al.* The ARVD/C genetic variants database: 2014 update. *Hum. Mutat.* **36**, 403–410 (2015).
  10. Richards, S. *et al.* Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
  11. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
  12. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. (2017). doi:10.1101/166298
  13. Linscheid, N. *et al.* Quantitative proteomics and single-nucleus transcriptomics of the sinus node elucidates the foundation of cardiac pacemaking. *Nat. Commun.* 1–19 (2019). doi:10.1038/s41467-019-10709-9
  14. Lundby, A. *et al.* Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics. *Nat. Methods* **11**, 868–874 (2014).
  15. Lundby, A. *et al.* In Vivo Phosphoproteomics Analysis Reveals the Cardiac Targets of  $\beta$ -Adrenergic Receptor Signaling. **6**, 1–14 (2013).

- Accepted Article
16. Romero, J., Mejia-Lopez, E., Manrique, C. & Lucariello, R. Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC/D): A Systematic Literature Review. *Clin. Med. Insights Cardiol.* 97 (2013). doi:10.4137/CMC.S10940
  17. Hall, C. L. *et al.* Frequency of genetic variants associated with arrhythmogenic right ventricular cardiomyopathy in the genome aggregation database. *Eur. J. Hum. Genet.* **2**, 1–7 (2018).
  18. Cruz, F. M. *et al.* Exercise triggers ARVC phenotype in mice expressing a disease-causing mutated version of human plakophilin-2. *J. Am. Coll. Cardiol.* **65**, 1438–1450 (2015).

### Figure Legends:

Figure 1: High resolution mass spectroscopy based cardiac proteomics data for ARVC genes – Protein abundance ratios for proteins encoded by ARVC genes is evaluated across cardiac chambers for (A+B) atria versus ventricles and for (C) right versus left ventricle. (D) The relative abundance in the right ventricle of proteins encoded by ARVC genes is shown as a rank-plot of iBAQ corrected MS intensities. Data points are based on cardiac biopsies from *Sus scrofa* (n=3).

Figure 2: ARVC associated genetic variants separated into ‘likely disease-causing’ (ARVC associated variants not found in gnomAD), ‘Questioned as monogenic disease-causing’ (ARVC associated variants found in gnomAD with prevalence below 1:2000), and ‘highly unlikely monogenic disease causing’ (ARVC variants found in gnomAD with prevalence above 1:2000). Shown in orange are variants CADD scored to be in the top 0.1% most deleterious variants. Shown in blue are variants below this threshold.

	Gene	# Reported ARVC Variants	# Variants also in gnomAD	# ARVC Variants not in gnomAD
High confidence ARVC genes	<i>PKP2</i>	230	57	173
	<i>DSP</i>	164	56	108
	<i>DSG2</i>	77	33	44
	<i>DSC2</i>	59	29	30
	<i>LMNA</i>	19	2	17
	<i>DES</i>	14	1	13
	<i>JUP</i>	18	18	0
Medium confidence ARVC genes	<i>RYR2</i>	9	6	3
	<i>TTN</i>	11	8	3
	<i>SCN5A</i>	4	1	3
	<i>PLN</i>	4	1	3
	<i>CTNNA3</i>	3	0	3
	<i>CDH2</i>	2	0	2
	<i>TMEM43</i>	11	10	1
	<i>TGFB3</i>	2	1	1
Low confidence ARVC genes	<i>OBSCN</i>	1	1	0
	<i>MYBPC3</i>	1	1	0
	<i>DMD</i>	1	1	0
Total		630	226	404

Table 1 – Summary of genetic variants reported to be associated with ARVC. The number of variants identified within individual genes associated with ARVC are listed. In addition we summarize the number of these variants also found within the gnomAD database. The association to ARVC for variants summarized in “# Variants also in gnomAD” are dubious as to whether they are monogenic causes of ARVC. The genes are sorted into three confidence categories, high, medium, and low, based on the number of remaining ARVC variants.

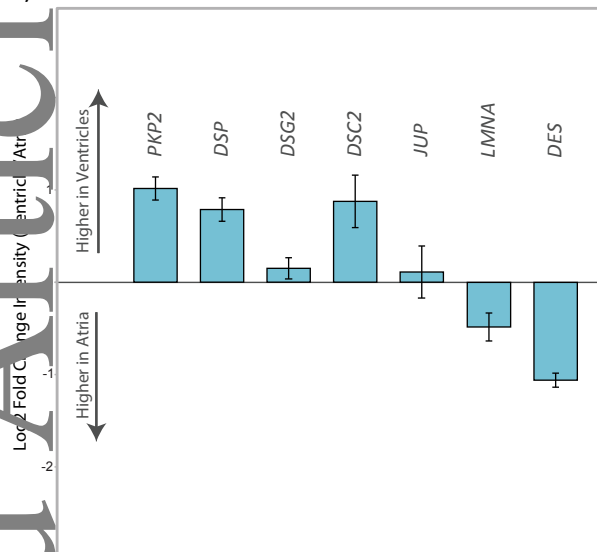
Gene	rsID	PheWeb AC	P-value	Beta	Odds-ratio	Phenotype
<i>LMNA</i>	rs142000963	2009				
<i>OBSCN</i>	rs200362121	653	0.027	1.6	5.0	Primary/intrinsic cardiomyopathies
<i>TTN</i>	rs72629793	219				
<i>TTN</i>	rs72646853	64	0.03	9.0	>5	Cardiac arrest
<i>TTN</i>	rs72650011	2174				
<i>TMEM43</i>	rs35028636	135	0.021	0.8	2.2	Congestive heart failure: nonhypertensive
<i>TMEM43</i>	rs145619906	1297				
<i>TMEM43</i>	rs201916031	619	0.025	1.7	>5	Primary/intrinsic cardiomyopathies
<i>TMEM43</i>	rs139590716	69				
<i>TMEM43</i>	rs113449357	11699				

DSP	rs121912998	764				
DSP	rs200473206	119	0.001	3.9	>5	Heart failure NOS
DSP	rs147315869	25	0.03	17.0	>5	Congestive heart failure?
DSP	rs28763961	130				
DSP	rs148147581	258				
DSP	rs150339369	182				
DSP	rs142494121	133				
DSP	rs141508330	142				
DSP	rs138907450	166	0.037	1.9	>5	Congestive heart failure
DSP	rs28763967	14312	0.011	1.9	>5	Other Cardiomyopathies
DSP	rs147415451	259				
DSP	rs116888866	148				
DSP	rs11558732	7532				
PKP2	rs139734328	536	0.022	12.0	>5	Other Cardiomyopathies
PKP2	rs146102241	3813				
PKP2	rs147240502	3221				Other
PKP2	rs200586695	123				
PKP2	rs143900944	23				
PKP2	rs139139859	1219				
PKP2	rs150821281	3508				
DSC2	rs147109895	777	0.01	2.0	>5	Cardiac arrest
DSC2	rs201856473	227	0.00031	15.0	>5	Ventricular fibrillation and flutter
DSC2	rs144799937	938				
DSG2	rs372174546	410	0.0005	40.0	>5	Other Cardiomyopathies
DSG2	rs191564916	728				
DSG2	rs139326669	111				
DSG2	rs201564919	59				

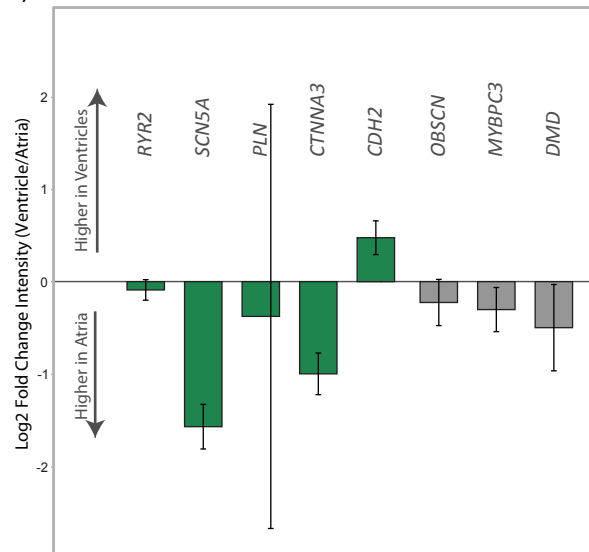
Table 2 - Evaluation of questionized ARVC variants in the phenome-wide association study of 409,000 participants in UK biobank. Questionized variants associated with ARVC, but also found in gnomAD, were evaluated using data from 409,000 participants in UK biobank using the PheWeb resource<sup>6</sup>. For relevant cardiovascular phenotypes, any cardiomyopathy, heart failure, ventricular fibrillation, or cardiac arrest, their p-value and odds-ratio were included.



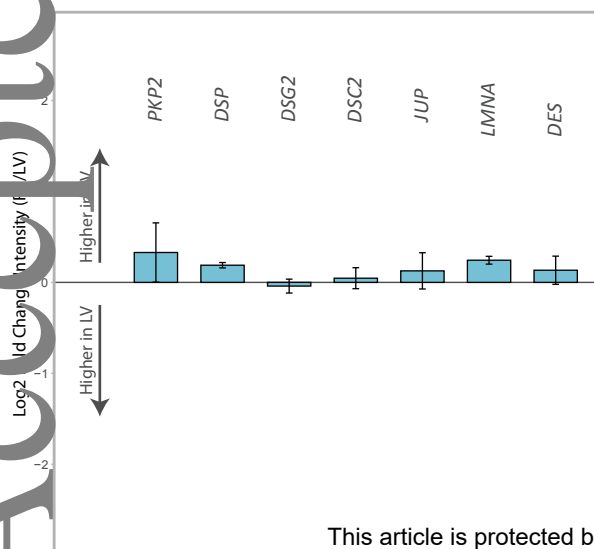
Ventricular/Atrial Protein Abundance Ratios of High Confidence ARVC Genes



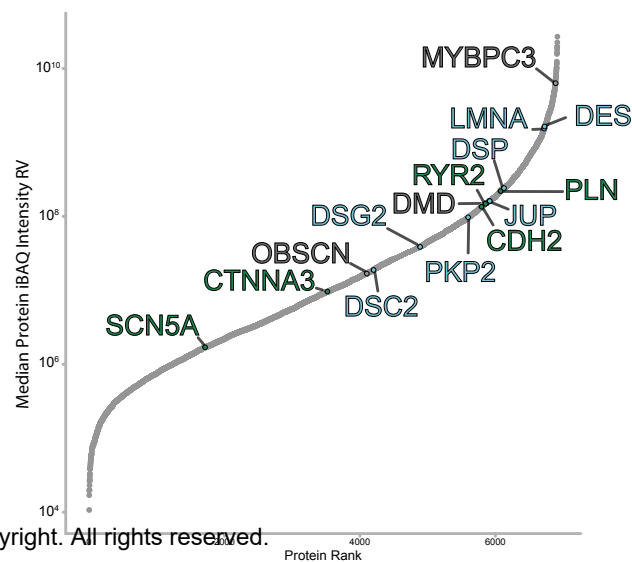
b) Ventricular/Atrial Protein Abundance Ratios of Non-high Confidence ARVC Genes



c) Right/Left Ventricular Protein Abundance Ratios of High Confidence ARVC Genes



d) Right Ventricle iBAQ Rank Plot



This article is protected by copyright. All rights reserved.

# CADD Scores of ARVC Associated Genetic Variants

Accepted Article

